



TITLE:

バンド幅問題に対するvolume
respecting embedding法の実験的
評価 (計算機科学基礎理論の新展開
)

AUTHOR(S):

駒原, 雄祐; 中野, 泰男; 山崎, 浩一

CITATION:

駒原, 雄祐 ...[et al]. バンド幅問題に対するvolume respecting embedding法の実験的評価
(計算機科学基礎理論の新展開). 数理解析研究所講究録 2003, 1325: 158-163

ISSUE DATE:

2003-05

URL:

<http://hdl.handle.net/2433/43190>

RIGHT:

バンド幅問題に対する volume respecting embedding 法の実験的評価

群馬大学・駒原 雄祐 (Yuusuke Komahara), 中野 泰男 (Yasuo Nakano), 山崎 浩一 (Koichi Yamazaki)

Department of Computer Science,
Gunma University

1. はじめに

1998 年に Feige が, バンド幅問題に対する近似率 $O((\log n)^3 \sqrt{\log n \log \log n})$ の近似アルゴリズムを提案した [8]. Feige のアルゴリズムが提案される以前では, Cuthill-McKee(CM) 法や改良逆 Cuthill-McKee(GPS) 法などの (バンド幅問題に対する) 幾つかの発見的手法が知られていた. しかしそれらアルゴリズムの近似率の正確な解析はされておらず, 実験的な評価のみであった. 文献 [8] で Feige は, 従来と全く異なる手法を提案し, さらに従来無し得なかった近似率の解析を初めて行った. 本論文では, バンド幅問題に対し, 良く知られていて比較の実装が簡単な Cuthill-McKee(CM) 法と Feige のアルゴリズムとの比較を行う.

1.1. バンド幅問題とは

$G = (V, E)$ をグラフとし, $LO(G)$ を V から $\{1, \dots, |V|\}$ への全単射の集合とする. グラフ G のバンド幅, $bw(G)$, は,

$$bw(G) = \min_{f \in LO(G)} \max_{\{u, v\} \in E} |f(u) - f(v)|$$

で定義される.

M を n 次の対称正方行列とする. M の全ての行に対して, 対角成分から非零成分までの列数の最大値, $bw(M)$, を行列 M のバンド幅と呼ぶ. PM_n をサイズ n の置換行列の集合とする.

$$b = \min_{P \in PM_n} bw(PMP^T)$$

のとき, 行列 M は最小バンド幅 b を持つ行列に変換可能であると言う. グラフのバンド幅を求める問題や, 行列を最小バンド幅を持つ行列に変換する問題をバンド幅問題と呼ぶ.

1.2. バンド幅問題の応用例

帯行列に対するガウスの消去法 M を n 次の対称正方行列とする. バンド幅問題を解くことにより, M をバンド幅 b を持つ行列に変換できたとする. このとき, (通常 $O(n^3)$ かかる) ガウスの消去法は $O(nb^2)$ で実行でき, $b \ll n$ の場合計算コストがかなり削減できる.

1.3. バンド幅問題の研究の歴史

1.3.1. レベル構造

バンド幅問題に対して従来からよく用いられている CM 法や GPS 法などの発見的手法は, レベル構造を構成している. グラフ $G = (V, E)$ の頂点の部分集合の列 (L_1, \dots, L_M) がレベル構造を成すとは,

1. $\sum_{i=1}^M L_i = V$,
2. $\{u, v\} \in E, u \in L_i, v \in L_j \Rightarrow |i - j| \leq 1$

を満たすときをいう. $\max_{1 \leq i < j \leq M} |L_i \cap L_j|$ をレベル構造 (L_1, \dots, L_M) の幅と呼ぶ. グラフ G のレベル構造での最小の幅を $lw(G)$ で表すと, $lw(G) \leq bw(G) \leq 2lw(G) - 1$ という関係が成り立つ. 従って, もし最小の幅を持つレベル構造を見つけることができれば, バンド幅の 2 倍近似が実現できることになる. しかしながら, 最小の幅を持つレベル構造を見つけることは NP 困難であることが知られている [2].

1.3.2. 距離に基づいたレベル構造

先に述べた通り, 最小の幅を持つレベル構造を見つけることは計算量的に困難である. 従って, 最小の幅は保証しないが, 比較的簡単にレベル構造が実現できる, 距離に基づいてレベルを構成する方法がよく用いられる.

R_0 を頂点のある部分集合とし, R_i を R_0 から距離 i にある点の集合とする. このとき R_0, R_1, \dots, R_h (h は R_0 から最も遠い点までの距離) はレベル構造を成す. 通常 R_0 のサイズは 1 である. 距離に基づいてレベルを構成する方法は手軽であるが, (理論的には) それにより得られるレベル構造の幅が最小の幅 $lw(G)$ よりかなり大きくなる場合が起こり得る. また距離に基づいて構成されたレベルの中での最小幅を求める問題は, たとえグラフを木に制限したとしても, $4/3$ 未満に近似率を下げられないことが知られている [17]. 実際, 距離に基づいて得られたレベル構造の幅が $O(lw(G) \log |V|)$ となる (人工的な) グラフは簡単に作ることが出来る. しかしながら, ランダムに生成されたグラフなどに対しては, この距離に基づいたレベル構造を用いたバンド幅問題の発見的手法は実験的には良い結果を出力することが知られている [14].

1.3.3. Cuthill-McKee(CM) 法

CM 法は Cuthill と McKee により 1960 年代後半に提案された方法で、距離に基づいたレベル構造を用いている [6]. CM 法の改良として GPS 法が良く知られている [9].

1. 各頂点 v に対し, $R_0 = \{v\}$ とし距離に基づくレベル構造を生成し, そのレベル構造の幅を計算する¹.
2. 得られた幅の中で最小の幅を持つレベル構造 R_0, \dots, R_h を選ぶ.
3. $u \in R_i, v \in R_j, i < j \Rightarrow f(u) < f(v)$ となるようにある規則にしたがって各頂点の並び (すなわち $f: V \rightarrow \{1, \dots, |V|\}$) を計算する².

1.3.4. バンド幅問題の難しさ

バンド幅問題は入力を最大次数が 3 の木に制限しても, 対応する判定問題が NP 完全である³ [10]. また, 入力を木に制限しても, 近似率は $4/3$ より良くならない [3]. 一般のグラフに対しては (つまり, 入力に何も制限を設けない場合), ($P \neq NP$ の仮定の下で) 定数近似は不可能であることも知られている [15]. 従って, Feige のアルゴリズムの近似率は決して悪くはないと言える. さらに, 与えられた正整数 k に対しバンド幅が k 以下であるかの判定問題は実質 brute force で行う方法しかないことが予想されている [4].⁴

1.3.5. 最近の研究

バンド幅問題に対する他の近似アルゴリズムとしては, Feige の論文 [8] と同年に, 半正定値計画法と射影法を組み合わせた技法が [1] で発表されている. Feige の論文 [8] 以降, [13, 7, 11] などの幾つかの論文が発表されている. 文献 [7] では Feige のアルゴリズムの近似率が $O((\log n)^3 \sqrt{\log \log n})$ に改善されている. グラフをあるノルム空間に埋め込む研究やランダム射影 [16] の研究が近年盛んに行われている.

¹ 計算時間を節約したい場合, 次数の低い頂点 v のみに対し $R_0 = \{v\}$ の計算を行う

² 規則の詳細な記述は文献 [5] を参照

³ 良く知られている問題で, 木に制限しても NP 完全であるものはあまり無い (他には例えば subgraph isomorphism や achromatic number など)

⁴ 正確には fixed parameterized intractable であることが知られている.

2. volume respecting embedding 法

Linial 等は, 各 2 頂点に対し, それらのグラフ上での距離をそれらの空間上での距離以下に保ち, かつあまり短くならないように, グラフをあるノルム空間に埋め込む研究を行った [12]. Feige はこの研究に触発され volume respecting embedding 法の研究を開始した. Feige は以下の性質を満たす (グラフの空間への) 埋め込みのアルゴリズムを開発した.

埋め込みが満たす性質 P

1. 任意の 2 頂点 u, v に関して, u と v の空間上での距離は高々グラフ上での距離,
2. 任意の k 頂点に対し, それらの k 頂点成す単体の $k-1$ 次元の体積が, ある程度の大きさを持つ.

この性質 2 の $k=2$ の場合は, 任意の 2 頂点成す 1 次元の体積つまり長さが小さくならないことを意味する. つまり, Linial 等の埋め込みは Feige の埋め込みの $k=2$ の場合に相当し, Feige の埋め込みは Linial 等の埋め込みの拡張と言える.

2.1. 基本アイデア

VRE 法は, 1) 埋め込み と 2) 射影 の 2 段階からなる. 埋め込みでは, 体積をある程度保つよう, グラフの各頂点を $L = \Theta((\log n)^4 \log \log n)$ 次元空間に埋め込む. 性質 1 を満たさなければならぬので, そのように埋め込むには工夫が必要である. 射影は空間内にランダムな直線を引き, その直線に, 埋め込まれた頂点が射影される. この直線上に射影された並びが出力解となる.

ランダムに引いた直線に各点を射影すると, 当然直線の最も端に射影された 2 つの頂点 (右端と左端) が存在する. この 2 頂点の距離が短いとバンド幅が大きくなる傾向がある. この 2 頂点の距離が短くなる確率は体積の大きさに反比例するので, 性質 2 が必要となる.

2.2. 近似率の算定方法

近似アルゴリズムでの近似率の計算には必ず最適解の下界 (最大化問題では上界) が関係している. この下界が良ければ, 近似率も良くなる傾向にある. Feige のアルゴリズムの近似率の解析では, local density ($ld(G)$)⁵ が下界として用いられている⁶. つまりは, Feige は [8] で, $bw(G) \leq ld(G)(\log |V(G)|)^3 \sqrt{\log n \log \log n}$ を示した.

⁵ $ld(G)$ 自身を求める問題は APX 完全であることが知られている

⁶ バンド幅問題では, この $ld(G)$ が下界として良く用いられる

2.3. アルゴリズム

アルゴリズムは1) 埋め込みと2) 射影の2段階からなる。どちらの段階も確率的アルゴリズムなので、良い解を得るためには適当な回数繰り返しが必要となる。射影よりも埋め込みの方が手間がかかる。時間計算量は $O(|E(G)|L) \simeq O(|E(G)|(\log n)^5)$ である。

埋め込みのアルゴリズム

入力: グラフ $G = (V, E)$

出力: $\varphi: V \rightarrow R^L$

1. $L = ((\log n)^4 \log \log n)$, $k_j = \lceil j \log n / L \rceil$, $p_j = 2^{-k_j}$, ($1 \leq j \leq L$) とする。
2. L 個の部分集合 S_j を作成する:
各頂点 v_i に対し, v_i を確率 p_j で集合 S_j に入れる。
3. 各頂点 v_i に対し, $\varphi(v_i)$ の各 j 座標 $\varphi_j(v_i)$ を $\varphi_j(v_i) = d(S_j, v_i)$ とセットする。

射影のアルゴリズム

入力: φ

出力: 頂点の並び

1. ランダムに $r_j \in N[0, 1]$ を選択する。
2. 各頂点 v_i に対し, $h(v_i) = \sum_{j=1}^L r_j \varphi_j(v_i)$ を計算する。
3. 頂点集合 V を $h(v_i)$ の値で昇順にソートし, その並びを出力する。

3. 実験結果 CM法 vs. VRE法

7回の埋め込み \times 40回の射影を1セットと勘定する。1セットの中で最も良い値を解として出力する。実験環境は, CPU, intel pentium 4 / 2.2GHz., メモリ, 1GB., OS, Vine Linux 2.5, コンパイラ, Java2SDK 1.4.0 である。

3.1. ランダムグラフに対する結果

それぞれ辺の密度が異なる 3000 頂点の 50 個のランダムグラフに対し実験を行った。 n を頂点数とすると, $L = ((\log n)^4 \log \log n)$ なので $n = 3000$ のとき $L = 62835$ となる。

1つのランダムグラフに対し, 幾つかの次元で VRE法を行った。正確には, $L/10$ 次元から始め, 次元が 20 以上の間は $1/2$ 倍づつして行き, 20 を下回った時点で 1 づつ減らし, 次元が 2 になるまで実験を行った。我々の実験環境では L 次元での VRE法が実行できなかった。そのため, $L/10$ 次元から実験を始めた。

これにより1つのランダムグラフに対し 19 個のバンド幅の値を得るが, その中での最も良いバンド幅を得る次元は全て一桁台であった。また,

次元が小さくなるに連れ出力解が良い値を取り, 最も良い値を出す次元は 3 から 8 の間に集中していることがわかった。典型的な例として, 辺の密度が異なる 3000 頂点の 3 つのランダムグラフでのバンド幅と次元の関係を図 1 に, 処理時間と次元の関係を図 2 に示す。

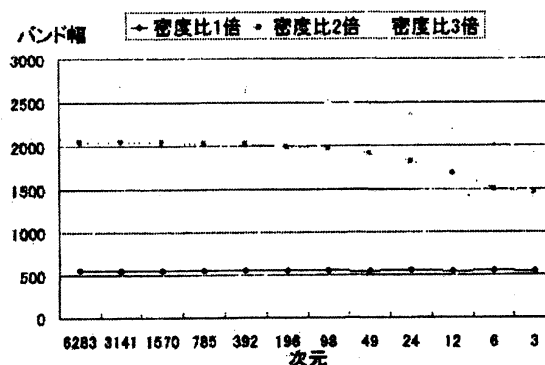


図 1: バンド幅と次元の関係

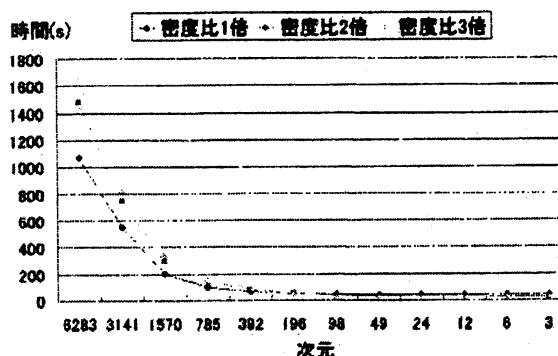


図 2: 処理時間と次元の関係

3.1.1. 解の精度での評価

大きい次元 ($L/10 = 6283$) と小さい次元 (5) の 2 つの次元での VRE法と CM法のそれぞれの出力解の値が図 3 にプロットされている。

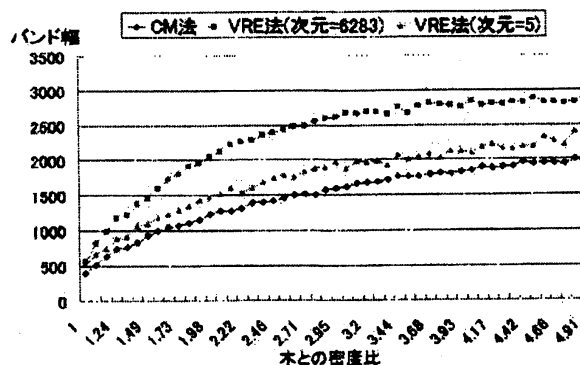


図 3: 3000 頂点のランダムグラフでの解の比較

ランダムグラフでの解の精度においては、図3が示すとおり、次元の大きさにかかわらず CM 法の圧勝である。

3.1.2. 処理時間での評価

大きい次元 ($L/10 = 6283$) と小さい次元 (5) の2つの次元での VRE 法と CM 法のそれぞれの処理時間が図4にプロットされている。

ランダムグラフの処理時間においては、図4が示すとおり、次元が大きい場合は CM 法の圧勝である。図4ではわかりにくいですが、次元が小さい場合は、辺の密度が小さいと CM 法の方が処理時間は短いですが、辺の密度が多くなると逆転が起きている。

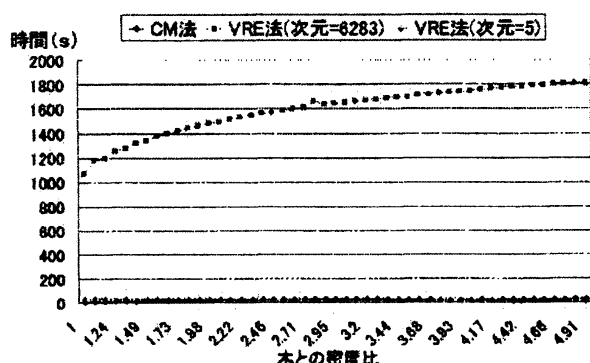


図4: 3000 頂点のランダムグラフでの処理時間の比較

3.2. CM 法が苦手とする特殊なグラフ

3.2.1. 風車グラフ

風車グラフとは図5にあるように、2部グラフの辺を根に近づく程より多くストレッチして得られるグラフである。図6が示すように、解は VRE 法がかなり良い。

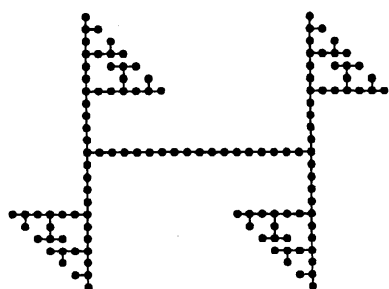


図5: 風車グラフ

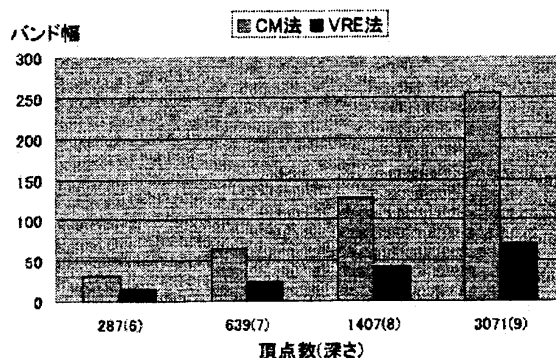


図6: 風車グラフ

図7は、深さ9(3071 頂点)の風車グラフに対しての、CM 法と VRE 法の処理時間の比較を示している。CM 法より良い値を出すかまたは1セット処理しきるまでの時間が VRE 法の曲線として記されている。

次元が小さい場合でも、VRE 法が CM 法より良い値を得るためには、少ない埋め込みで十分であることが図7から読み取れる。次元が2で VRE 法の曲線が跳ね上がっているのは、埋め込みの失敗が続いたためと考えられる。

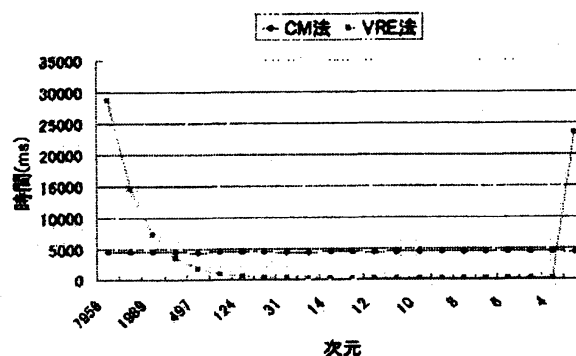
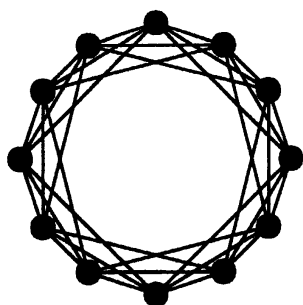
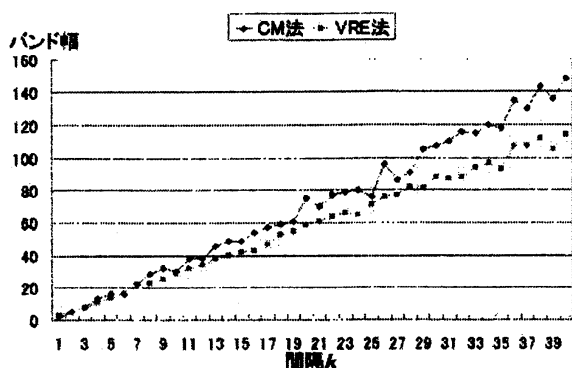


図7: 風車グラフでの時間の比較

3.2.2. C_n の k 乗グラフ

C_n の k 乗グラフとは、 n 頂点から成るサイクル C_n の各頂点 v に対し、 v と距離 k 以下にあるすべての頂点を辺で結ぶことにより得られるグラフである。

辺を結ぶ間隔 k が大きくなる程 CM 法と VRE 法の差が開いて来ることが図9からわかる。

図 8: C_{12} の 3 乗グラフ図 9: C_{1000} の k 乗グラフでの解の比較

4. 近似率での評価 (完全 2 分木に対する結果)

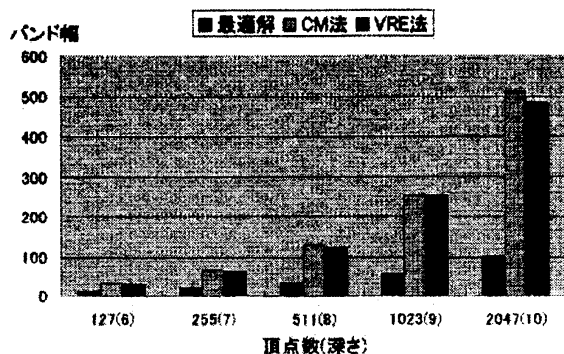


図 10: 2 部グラフでの解の比較

図 10 からわかるように解の精度に関しては CM 法と VRE 法の差はあまり無い。約 2000 頂点で両者の解は最適解の約 5 倍程度で抑えられている。

5. まとめ及び考察

5.1. VRE 法の評価

文献 [8] に沿って、

1. 埋め込み先の次元を大きくとり、
2. 埋め込みのアルゴリズムを繰り返し、
3. 射影のアルゴリズムも繰り返す

と処理時間は、従来手法と比べるとかなりかかる。また解の精度も CM 法と比べやや劣る。ランダムなグラフに対しては、時間及び精度の両面で VRE 法に勝ち目は無いという印象を (実験結果から) 受ける。しかしながら、CM 法が苦手とする風車グラフや C_n の k 乗グラフに対しては、VRE 法の方が随分良い値を出力する。

埋め込みのアルゴリズムのステップ 3 の座標の計算は、幅優先探索により行われている。次元の大きさ分幅優先探索が行われ、これが処理時間がかかる原因である。これらのことは、文献 [8] の結論にかかれている内容と一致している。実装という点から言えば、VRE 法は CM 法より手軽である。

5.2. 次元の大きさの設定

埋め込み先の空間の次元の値が小さくなればなる程、処理時間が格段に短くなるという利点がある。

今回の実験の結果は、次元はそれほど大きくする必要が無く、むしろ小さく抑えた方が解の精度が上がる事を示している。この現象はランダムグラフに限らず、風車グラフや C_n の k 乗グラフなどの特殊なグラフに対しても見られる。文献 [8] では、

1. 次元が $\Theta((\log n)^4 \log \log n)$ 程度あれば埋め込みのアルゴリズムが高い確率で埋め込みの性質 P を満たす、
2. 性質 P を満たす埋め込みに対して射影のアルゴリズムを適用すると、比較的精度の良い解が得られる、

が証明されている。今回の実験から導かれる推測として:

1. 次元が小さくても性質 P を満たすことができる、
2. 次元が小さい場合は性質 P を満たさないが、別の性質 P' を満たし、 P' を満たす埋め込みに対して射影のアルゴリズムを適用すると、比較的精度の良い解が得られる、
3. どんな埋め込みに対しても、射影のアルゴリズムを適用すると比較的精度の良い解が得られる、

の 3 つが考えられる。しかし、小さい次元でランダムな埋め込みを行い、その後射影のアルゴリズムを適用すると極めて悪い精度の解しか得られない事が実験によりわかるので、可能性としては 1 か 2 のどちらかだと考えられる。つまり、小さい次元での埋め込みのアルゴリズムは、バンド幅

問題に有利に働く何らかの性質を満たすよう埋め込みを行っていると考えられる。しかしその性質が P かそれ以外の性質なのかは、現時点では不明である。

6. 今後の課題

次元の大きさの設定

精度の良いバンド幅を求めるためには、実験的には、小さい次元で良い。(もし事実がそうであれば) 比較的良好な解が得られる理由の解明。

Worst case analysis

従来ある多くの発見的手法に対して、その手法ではうまく働かない人工的なグラフを比較的に簡単に作る事が出来る。VRE 法ではうまく働かない(つまり解の精度が近似率程度に悪くなる)グラフを求める。

謝辞 本研究は、文部省科学研究費補助金(課題番号 13780196)の助成を受け行ったものである。

参考文献

- [1] A.Blum, G.Konjevod, R.Ravi and S.Vempala, Semi-Definite Relaxations for Minimum Bandwidth and other Vertex-Ordering problems, Proc. of 30th STOC, (1998), 100-105.
- [2] H.L.Bodlaender, The complexity of finding uniform emulations on paths and ring networks, Information and Computation 86 (1990) 87-106.
- [3] G.Blache and M.Karpinski, On Approximation Intractability of the Bandwidth Problem, ECCC TR98-014 (1998).
- [4] H.L.Bodlaender, M.R.Fellows, and M.T.Hallett, Beyond NP-completeness for problems of bounded width (extended abstract): hardness for the W hierarchy, Proc. of 26th STOC, (1994), 449-458.
- [5] P.Z.Chinn, J.Chvatalova, A.K.Dewdney, and N.E.Gibbs, The bandwidth problem for graphs and matrices—a survey, J. Graph Theory 6 (1982), 223-254.
- [6] Cuthill, E., McKee, J.: Reducing the bandwidth of sparse symmetric matrices. In: Proc. ACM Nat.Conf. New York (1969), 157-172.
- [7] J.Dunagan and S.Vempala, On Euclidean embeddings and bandwidth minimization, Proc. of the 5th Workshop on Randomization and Approximation, (2001).
- [8] U.Feige, Approximating the bandwidth via volume respecting embeddings, (Extended abstract) STOC (1998) 90-99: (Journal version) Journal of Computer and System Sciences 60(3), (2000), 510-539.
- [9] N. E. Gibbs, W. G. Poole, Jr., and P. K. Stockmeyer, An algorithm for reducing the bandwidth and profile of a sparse matrix, SIAM J. Numer. Anal. (1976), 13:236-250.
- [10] M.Garey, R.Graham, D.Johnson, and D.Knuth, Complexity Results For Bandwidth Minimization, SIAM J. Appl. Math. 34 (1978), 477-495.
- [11] A.Gupta, Improved Bandwidth Approximation for Trees and Chordal Graphs, J. Algorithms 40(1) (2001), 24-36.
- [12] N.Linial, E.London, Y.Rabinovich, The Geometry of Graphs and Some of its Algorithmic Applications, Combinatorica 15(2) (1995), 215-245.
- [13] S.Rao, Small distortion and volume preserving embeddings for planar and Euclidean metrics, Annual Symposium on Computational Geometry (1999), 300-306.
- [14] J.S.Turner, On the Probable Performance of Heuristics for Bandwidth Minimization, SIAM J. Comput. 15(2) (1986), 561-580.
- [15] W.Unger, The Complexity of the Approximation of the Bandwidth Problem, Proc. 39th FOCS (1998), 82-91.
- [16] S.Vempala, Random Projection: A New Approach to VLSI Layout, FOCS (1998).
- [17] K.Yamazaki, On Approximation Intractability of the Path-Distance-Width Problem, Discrete Applied Mathematics Vol.110 (2001), 317-325.